

**Original citation:**

Quareshy, Mussa, Prusińska, Justyna, Li, Jun and Napier, R. (Richard). (2017) A cheminformatics review of auxins as herbicides. Journal of Experimental Botany.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/92069>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

This is a pre-copyedited, author-produced PDF of an article accepted for publication in Journal of Experimental Botany following peer review. The version of record Quareshy, Mussa, Prusińska, Justyna, Li, Jun and Napier, R. (Richard). (2017) A cheminformatics review of auxins as herbicides. Journal of Experimental Botany is available online at:  
<http://dx.doi.org/10.1093/jxb/erx258>

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

**A cheminformatics review of auxins as herbicides**

**Mussa Quareshy<sup>1,3</sup>, Justyna Prusinska<sup>1</sup>, Jun Li<sup>1,2</sup>, Richard Napier<sup>1,3</sup>**

<sup>1</sup> School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK

<sup>2</sup> Department of Pesticide Science, College of Crop Protection, Nanjing Agricultural University, Weigang 1, Nanjing, Jiangsu Province, P.R. China

<sup>3</sup> Corresponding authors:

[m.quareshy@warwick.ac.uk](mailto:m.quareshy@warwick.ac.uk); [Richard.napier@warwick.ac.uk](mailto:Richard.napier@warwick.ac.uk)

**Highlight**

We review cheminformatics of herbicides with a special focus on auxins and highlight the utility of DataWarrior for analysing physicochemical properties in the context of herbicide discovery.

**Keywords:** physicochemical, compound, chemistry, agriculture, agrochemical, molecule, biochemistry, herbicide resistance, mode of action.

## Abstract

Herbicides are an important asset in ensuring food security, especially when faced with ever-increasing demand on food production to feed the planet. The current selection of herbicides is increasingly encountering resistance in the agricultural weeds they once targeted effectively. It is imperative that new compounds, or even better, new modes of actions are discovered in order to overcome such resistances. This cheminformatics review looks at the current herbicides and evaluates their physiochemical properties on a class by class basis. We focus in particular on the synthetic auxin herbicides, Herbicide Resistance Action Committee class O, analysing these against herbicides generally, and for class-specific features such as mobility in plant vasculature. We summarise the physiocochemical properties of all the 24 compounds used commercially as auxins, and relate these results to ongoing approaches to novel auxin discovery. We introduce an interactive, open source cheminformatics tool known as Data Warrior to herbicide discovery, complete with records for over 300 herbicidal compounds, and hope this helps researchers as part of a rational approach to not only auxin discovery, but agrochemical discovery generally.

## Definitions:

**Log P:** Logarithm of the partition coefficient of the compound between an organic phase (octanol) and an aqueous phase (water) at a pH where all of the compound molecules are in the neutral form.

$$\text{Log P} = \log([\text{Compound}_{\text{organic}}]/[\text{Compound}_{\text{aqueous}}])$$

**Log D:** Logarithm of the distribution coefficient of the compound between an organic phase (octanol) and an aqueous phase (e.g., buffer) at a specified pH (x). A portion of the compound molecules may be in the ionic form and a portion may be in the neutral form.

$$\text{Log D}_{\text{pHx}} = \log([\text{Compound}_{\text{organic}}]/[\text{Compound}_{\text{aqueous}}])$$

**pKa** – The acid dissociation ( $K_a$ ) constant is the measure of the strength of an acid in an aqueous solution, and pKa is the negative log of the acid dissociation constant  $K_a$ .



$$\text{pKa} = -\log([\text{H}^+] \times [\text{A}^-] / \text{HA})$$

**cLogS** - Solubility is the maximum dissolved concentration under given solution conditions. (This is a calculated value as opposed to an experimentally derived figure).

$$\text{Log S} = 0.8 - \text{Log P} - 0.01(\text{MP}-25)$$

Here S is solubility, Log P is the octanol/water partition coefficient, and MP is the melting point (a measure of crystal lattice strength).

## Introduction

Herbicides are of considerable economic importance in many areas of agriculture, horticulture, turf and landscape management, and contribute to global food security. Weeds reduce crop yields significantly as they compete for light, water and nutritional resources. There have been reports of yield losses ranging from 30% to high 90%'s, varying for crops and locations. One such extreme case saw a 92% loss in Venezuela (1987) for cassava crops (Cobb and Reade 2010), and even though losses are generally lower, weeds are responsible on average for between 8 - 13% of all losses in global crop yields annually, and weeds were by far the largest contributor to crop losses (34%), ahead of animal pests (18%) and pathogens (16%) between 2001-2003 (Oerke 2006).

88 More recent surveys by the Weed Science Society of America have estimated the  
89 average annual value of losses to weeds in corn crops in the USA alone to be \$27  
90 billion, and for soybean \$16 billion from surveys between 2007 and 2013 (Soltani et al.  
91 2016).

92  
93 There are 17 classes of herbicide according to the HRAC (Herbicide Resistance Action  
94 Committee) classification (Délye, Jasieniuk, and Le Corre 2013); Supplementary Table  
95 1). These classes represent different target sites and/or modes of action (i.e. the ways  
96 in which the herbicide controls susceptible plants). For example, herbicides that act as  
97 auxins are class O (Figure 1), and compounds that impair auxin transport are class P.  
98 Synthetic auxins mimic the effect of indole-3-acetic acid (IAA), the principal natural  
99 auxin in higher plants. They bind to Transport Inhibitor Response 1 (TIR1) (Dharmasiri  
100 et al., 2005a; Kepinski and Leyser, 2005) and Auxin F-Box (AFB) auxin receptors  
101 (Dharmasiri et al., 2005b).

102  
103 Auxin perception has been covered in the literature in some detail (Mockaitis and  
104 Estelle 2008; Sauer et al., 2013; Tan et al. 2007), but briefly, auxin responses are  
105 mediated through an Skp (S-phase Kinase-associated protein 1), Cullin, F-box protein  
106 (SCF) complex which is an E3 ubiquitin ligase. TIR1 is the F-box protein linked via  
107 Skp1 (also called ASK1) to Cullin 1 (CUL1), which is known to be the scaffold protein  
108 of SCF complex. At low auxin concentrations, Aux/IAA transcriptional repressor  
109 proteins repress genes targeted by the Auxin Response Factor (ARF) transcriptional  
110 activators. As auxin concentrations rise, auxin binds to TIR1 creating a binding site for  
111 the Aux/IAA proteins on the SCF<sup>TIR1</sup> complex. The Aux/IAA proteins are then  
112 ubiquitinated and broken down, releasing gene expression and stimulating growth and  
113 development. However, at high doses the synthetic auxins are phytotoxic, inducing  
114 widespread over-reaction to auxin stimulation which leads to injury and death  
115 (Grossmann, 2010). In part, phytotoxicity arises because the plants cannot readily  
116 control the concentration of synthetic auxins as readily as they can the concentration of  
117 natural auxin levels.

118  
119 There have been examples of resistance reported against all herbicide modes of action  
120 (Délye et al. 2013) with a recent review on global herbicide resistance reporting a total  
121 of 404 unique cases of herbicide-resistant weeds, with a further 11 cases reported a

122 year, on average (Heap 2014). It can take as few as six years of using herbicides with  
123 the same site of action for a resistance problem to appear. There have been 32 species  
124 with reported resistance to 2,4-D between 1957 and 2016 (Goggin et al., 2016). Some  
125 weeds have developed “cross resistance” and are resistant to multiple herbicides with a  
126 single mode of action such as 2,4-D and dicamba (Ghanizadeh and Harrington 2016;  
127 Schulz and Segobye 2016). Others show “multiple resistance” and are resistant to  
128 herbicides from two or more modes of action. Such multiple resistance has been  
129 reported for synthetic auxin herbicides for example wild radish (*Raphanus*  
130 *raphanistrum*) demonstrates resistance to ALS inhibitors (**Group B**), carotenoid  
131 biosynthesis inhibitors (**Group F**), EPSP synthase inhibitors (**Group G**) and synthetic  
132 auxins (**Group O**) (<http://www.weedscience.com/> - 2017).

133  
134  
135 If we are to achieve and maintain food security, there remains a need to discover novel  
136 compounds from existing herbicidal classes, as well as an outstanding need for the  
137 discovery of novel modes of action. The most recently introduced mode of action was  
138 the 4-Hydroxyphenylpyruvate dioxygenase (4-HPPD) class of inhibitors targeting  
139 photosynthesis, and this was in the late 1980’s (Duke 2012). In the case of auxin, two  
140 novel compounds in a new sub-class known as the 6-arylpicolinates have reached the  
141 marketplace over the last two years (Epp et al., 2016). This shows that auxin discovery  
142 remains an open book, and despite more than 70 years of exploration the possibility of  
143 yet more auxins has not been saturated. Research on the plant auxin receptors has led  
144 us and others to describe in detail the molecular binding site for these compounds (   
145 Calderon Villabobos et al., 2012; Lee et al. 2014; Uzunova et al. 2016) and to bring  
146 these two topics together we used this article to review the cheminformatics literature.  
147 We have applied state of the art cheminformatics tools to herbicides in general and this  
148 has allowed us to discuss the properties and design features of auxins in particular. Two  
149 earlier reviews of herbicide properties in general (Gandy et al. 2015; Tice 2001) have  
150 been useful reference points, although in this study we have refined our compound  
151 inputs to focus only on the bioactive versions of each compound, excluding the esters  
152 and other frequent conjugates used in formulations. We discuss some class-specific  
153 differences and compare in detail the cheminformatic characteristics of herbicide class  
154 O, the auxins.

155

## 156 Cheminformatics

157 Cheminformatics is a general term for the discipline of analysing and comparing the  
158 physicochemical properties of compounds *in silico*. Previous cheminformatics reviews  
159 of herbicides have evaluated similarities and differences between the available  
160 physicochemical properties of agro pharmaceuticals, including herbicides.  
161 Synonymous to this type of analysis is the set of parameters summarized in Lipinski's  
162 rule of 5 (Lipinski et al. 1997) which are designed to be pertinent to human "absorption,  
163 distribution, metabolism, and excretion" (ADME) pharmacokinetics. The rules state  
164 that orally bioavailable drugs will not contravene more than one of the following rules:  
165 a molecular mass of less than 500 daltons, no more than 5 hydrogen bond donors, no  
166 more than 10 hydrogen acceptors and an octanol-water partition coefficient (LogP)  
167 greater than 5. Whilst these "rules" have proved instructive, it is obvious that the target  
168 organisms and routes of administration are very different for herbicides and drugs.

169  
170 Previous work seeking Lipinski-like "agrochemical" rules did focus on herbicides (Tice  
171 2001), and showed that fewer hydrogen bond donors were allowed. Follow-up work  
172 presented a quantitative method to estimate pesticide likeliness for herbicides,  
173 insecticides and fungicides (Avram et al. 2014; Gandy et al. 2015), and Gandy et al.  
174 provided a useful open source database, with pre-calculated herbicide physiochemical  
175 properties. The database was available as an MS Excel® spreadsheet with added  
176 functionality to plot physiochemical data, to interrogate the database for comparisons,  
177 and the facility for updates, as necessary. Whilst these studies and resources have all  
178 provided useful insights for researchers in agrochemical discovery, we wondered  
179 whether or not these generalized approaches, even including just herbicides, may be  
180 too restrictive. Different physicochemical properties will be desirable depending on  
181 the herbicide mode and/or site of action.

182  
183 Past analyses have considered some herbicides in their commercially available  
184 conjugated forms, most frequently as esters. Whilst these compounds will eventually  
185 be metabolized *in vivo* into their active forms, the physiochemical analysis of these  
186 conjugates will not give a realistic representation of the active compound (frequently  
187 the free carboxylic acid). Ester derivatives are designed for compound stability and  
188 handling in the wider context of formulation, which is in contrast to drug discovery  
189 programs, which generally focus on active compounds. Drug delivery and formulation

190 is a later stage in development, even though they will also require ADME evaluation *in*  
191 *vitro* and *in vivo* (Knowles 1998). Additionally, despite the Gandy database resource  
192 searching through a flat-file, text databases do not allow researchers to search by  
193 chemical structure features. They provide chemical structures as images but, with 334  
194 compounds, the challenge to find a particular scaffold or functional group becomes very  
195 difficult for even the most chemically-astute researcher.

196

197 In this review, we offer a focussed and rational cheminformatics analysis of herbicides.  
198 We use a class-by-class approach, rather than a one-size-fits all approach. We highlight  
199 the limitations of studying compounds in their pro-drug form rather than active forms.  
200 We provide a collection of chemical structures in .sdf format and recommend the use  
201 of DataWarrior from openmolecules.org as a database platform. This is a stand-alone  
202 open source chemical viewing and cheminformatics tool, and it is compatible with all  
203 3 major operating system environments (Windows, Mac OSX and Linux). DataWarrior  
204 is able to assimilate chemical structures and related properties into a single file with  
205 many cheminformatics search facilities, as well as statistical and plotting functions.

206

#### 207 **A Database in DataWarrior**

208 Compound structures and their corresponding HRAC classes were added into  
209 DataWarrior [openmolecules.org] manually, removing ester groups and salts. Entries  
210 were saved in the .sdf file format. We used the cxcalc (ChemAxon; Supplementary  
211 Information 2) command line tool to calculate physiochemical properties for each  
212 compound (.sdf input and output files), appending the relevant physiochemical  
213 properties to the database. Parameters were calculated at pH 7.4, where relevant. Total  
214 surface area, polar surface area and relative polar surface area were calculated in  
215 DataWarrior. Data analysis and visualisation were performed in DataWarrior and  
216 Graphpad Prism v7.0a

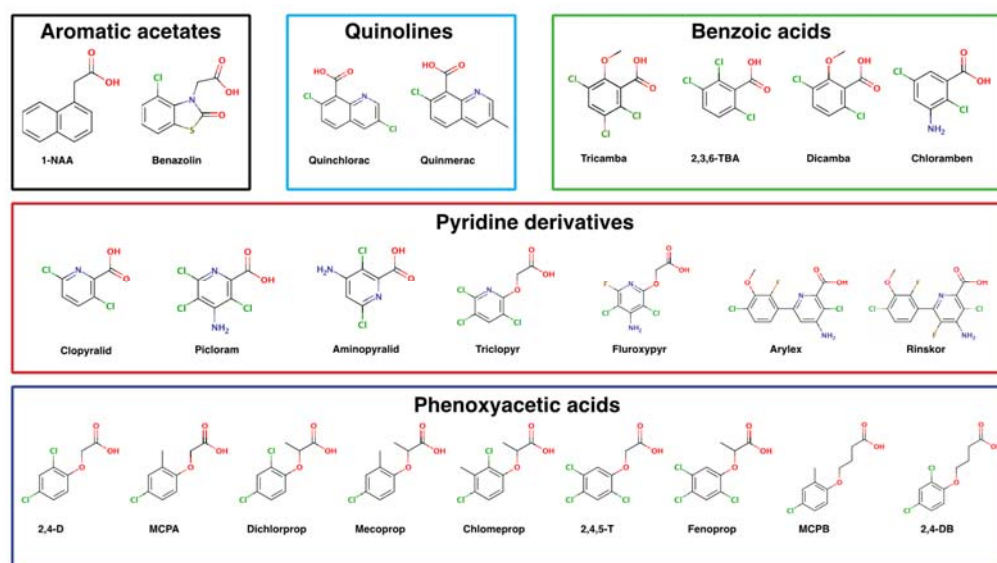
217

#### 218 **Features available in DataWarrior:**

219 There are a multitude of features offered in DataWarrior, most of which are beyond the  
220 scope of this work, and we encourage readers to use the software to gain familiarity.  
221 Datawarrior enables ready handling and organisation of chemical structures along with  
222 associated data. Viewing configurations may be customised and there are multiple tabs  
223 to switch between graphical data views or e.g. view structures only. Structures can be



entered manually or entire collections imported from .sdf files. We routinely view the database as a table (Supplementary Figure S1), showing structures with names and HRAC classes, for example (Supplementary Figure S2). Physiochemical properties can be calculated using the built-in physiochemical calculator tool (Supplementary Figure S3), whilst properties calculated elsewhere can be imported using corresponding compound ID's as unique keys in text files or chemical structures in .sdf format. Users can interrogate the database by keyword searches using the filter tools on the right side of the screen, often with additional interaction using intuitive sliders. The ability to search the database using chemical structures is a noteworthy feature (Supplementary Figure S4 shows results from a search entry to identify all compounds possessing a benzoic moiety). The interactivity also extends to generating graphical 2- and 3-dimensional data summaries by choosing data to plot with simple menu selections of bar charts (Supplementary Figure S5) or scatter plots (Supplementary Figure S6).

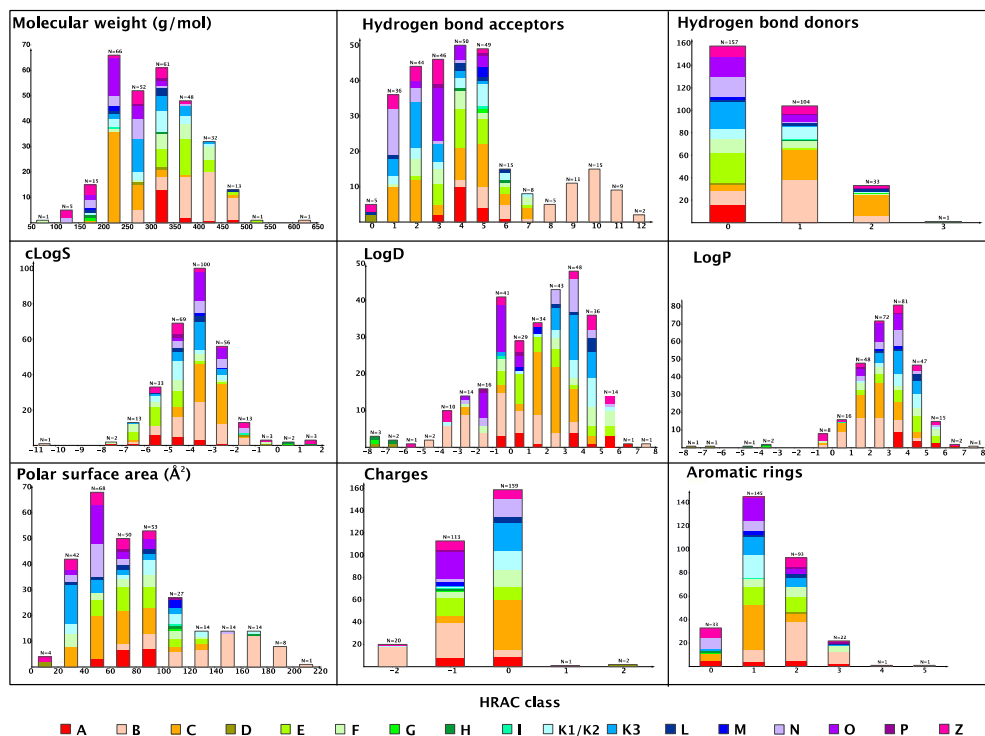


**Figure 1. Current synthetic auxin herbicides.** HRAC class O, grouped by their chemical types. (1-NAA = 1-Naphthaleneacetic acid, 2,3,6-TBA = 2,3,6-trichlorobenzoic acid, 2,4-D = 2-(2,4-dichlorophenoxy)acetic acid, MCPA = 2-(4-chloro-2-methylphenoxy)acetic acid, 2,4,5-T = 2-(2,4,5-trichlorophenoxy)acetic acid, MCPB = 4-(4-chloro-2-methylphenoxy)butanoic acid and 2,4-DB = 4-(2,4-dichlorophenoxy)butanoic acid.

## Trends observed

Six out of the 17 herbicide classes contained between 1 to 6 compounds, whilst the remaining classes had at least 19 or more compounds up to a maximum of 56. There

are 24 synthetic auxin compounds in class O (Figure 1). Distribution summaries of all classes for various physiochemical properties are shown in Figure 2. We observed no obvious class-specific trends for parameters such as hydrogen bond donor count. Most compounds had between 0 and 2 hydrogen bond donors, with only 1 compound having 3. Most compounds were either neutral or mono-anionic with a small number of compounds (mainly from Group B) being di-anionic. Similarly, most compounds had 1 or 2 aromatic rings. The aqueous solubility parameter cLogS showed a normal distribution for each class, as well as for individual classes, albeit over characteristic ranges.



**Figure 2. Physicochemical properties of herbicides.**

The distribution for 9 physiochemical properties with each herbicide class coloured differently (listed below the x-axis). Molecular weight, cLogS, LogD, LogP and polar surface areas are presented as binned groups on the x-axis, whilst H-Bond donor count, H-Bond acceptor count, charges and aromatic rings are shown as specified values.

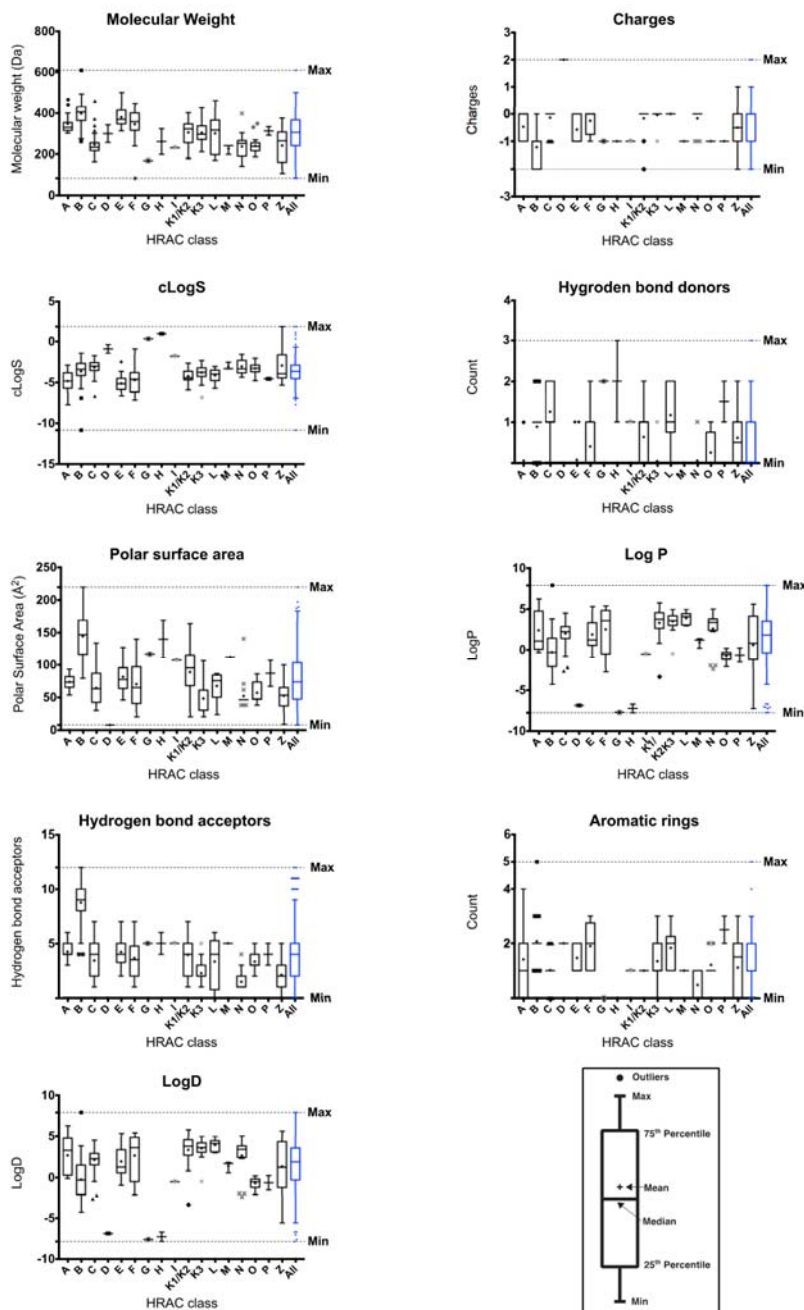
It is useful to compare these summaries for herbicides to Lipinski's Rules, and the more recent global analysis of these and other parameters such as "chemical beauty", a concept of desirability for physicochemical properties based on quantitative estimates of drug-likeness (Bickerton et al., 2012). Almost all herbicides have a molecular weight well below 500 (median 305); Figure 3; Gandy et al., 2015). This is common with drugs, where the median molecular weight is also just above 300 (Bickerton et al.

271 2012). The median number of hydrogen bond donors for drugs (aprox 1.5) is  
272 considerably higher than for herbicides (around 0), and there is a wide spread in the  
273 number of hydrogen bond acceptors for both families, each with a median around 4.  
274 The partitioning properties of drugs, represented by logP, has a broad distribution  
275 peaking between values of 1 and 4, with a median of approximately 3 (Bickerton et al.  
276 2012) whereas for herbicides as free active compounds this value is 2 (Figure 3).  
277 However, this is very class dependent and when the data are considered for the esters  
278 the median is similar to that for drugs at close to 4 (Gandy et al. 2015). Global  
279 comparisons against all insecticides and fungicides are presented in Avram et al. 2014.  
280

281 A comparative statistical breakdown for each parameter by class, and compared to all  
282 herbicides collectively is presented in Figure 2, with Table 1 showing a summary of a  
283 one-way ANOVA. These analyses of physicochemical properties for each herbicide  
284 class give a far more detailed picture than the generic plots of previous reviews  
285 (although in each case some specific breakdowns were presented; (Avram et al. 2014;  
286 Gandy et al. 2015; Tice 2001). For example, the majority of class B compounds (ALS  
287 inhibitors) are larger than 300 Da and display a comparatively high number of hydrogen  
288 bond acceptors (median close to 9 against a median of fewer than 5 for all other classes),  
289 with a correspondingly high polar surface area. Compounds in classes K3 and N  
290 (inhibitors of fatty acid synthase and elongase, respectively) have fewer hydrogen bond  
291 acceptors than average, again, associated with lower polar surface areas.

292

293 Considering the auxins, for lipophilicity (LogD and logP), the auxins cluster  
294 distinctively tightly, with values between 0 and 3 (small net lipophilicity). Class B is  
295 the only other large group with a mean close to 0, although Classes I (inhibitor of  
296 dihydropteroate synthase, one compound) and P (the auxin transport inhibitors, 2  
297 compounds) also share the same space. Auxins invariably have a formal charge of -1,  
298 representing ionisation of the carboxylic acid. The auxins are generally at the small  
299 (low molecular weight) end of the size spectrum, although the two new 6-  
300 arylpicolinate are clear outliers in the group for this property (Figure 3).



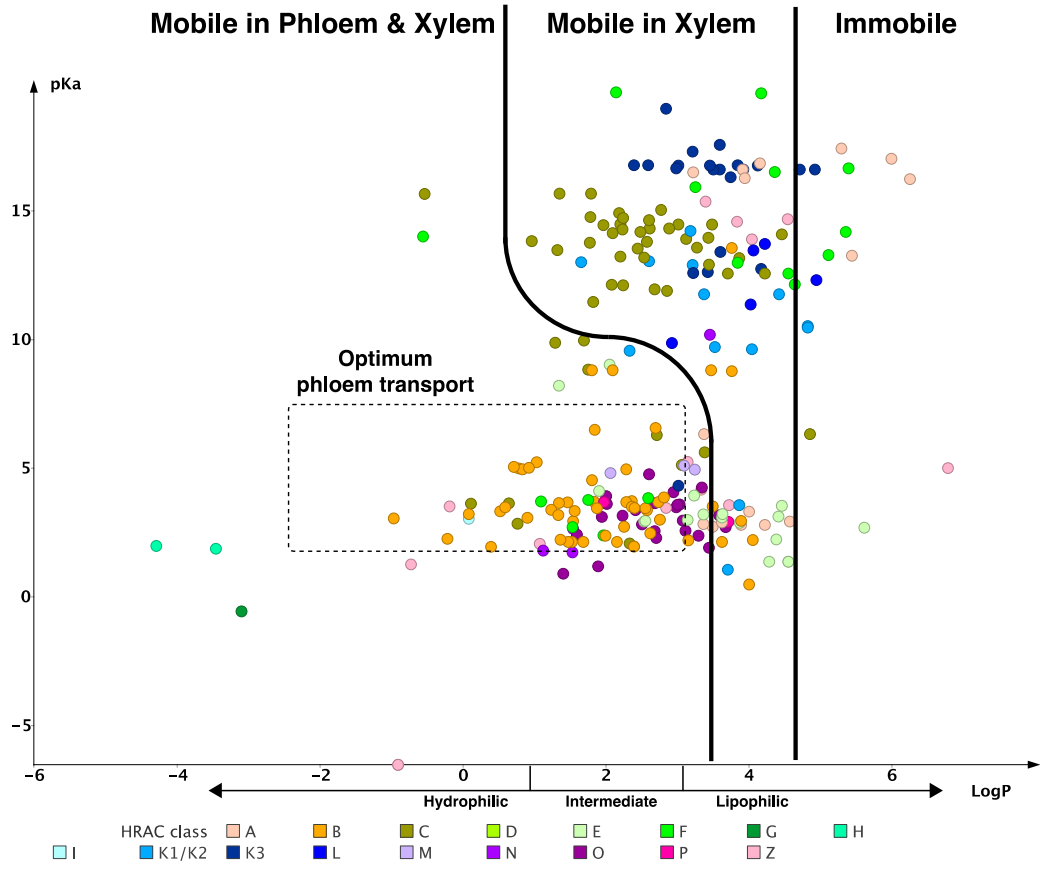
**Figure 3 Separating out physicochemical properties by class.**

A selection of physiochemical properties depicted as box plots (black) showing their group-specific trends compared to that of the complete dataset (shown in blue at the right)

#### Properties and functionality:

The relationship between LogP vs. pKa has been reported as a useful predictor of

311 compound mobility in plant vasculature (Bromilow et al., 1986; re-printed by Cobb and  
 312 Reade 2010). Most herbicides are predicted to be mobile in the plant and the Bromilow  
 313 mobility predictions are shown (Figure 4) overlaid onto values from our herbicide  
 314 database. 243 out of 295 compounds in our database are shown, as the remainder do  
 315 not possess acidic ionisable groups (hence no pKa). It can be seen that most herbicides  
 316 are predicted to be xylem and phloem mobile, with a few just outside the proposed  
 317 allowable region (Figure 4).

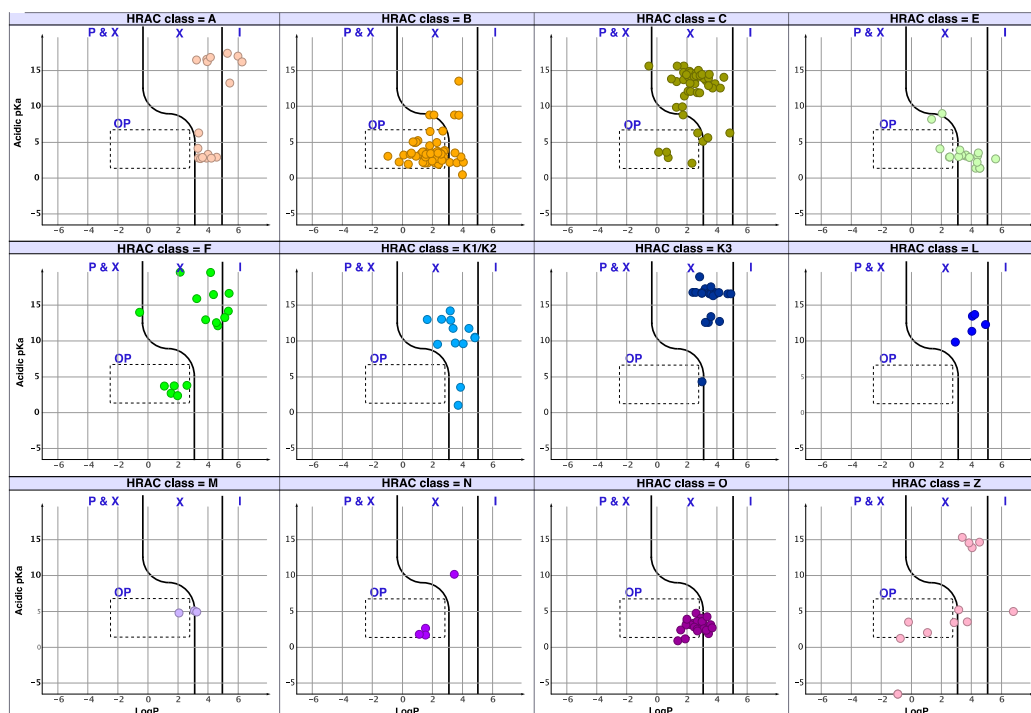


318

319 **Figure 4. Herbicide mobility in planta.**

320 The ranges limiting vascular mobility in phloem and xylem using the criteria defined  
 321 by Bromilow, Chamberlain, and Briggs 1986. The majority of herbicide actives lie  
 322 within the proposed ranges for mobility.  
 323

324 Interestingly, we observed some class-specific clusterings for the LogP vs. pKa analysis  
 325 (Figure 5), with classes C (Photosystem II inhibitors), K1/K2 (microtubule  
 326 depolymerisation) and K3 showing close clustering for mobility in the xylem. Classes  
 327 B, E (inhibitors of protoporphyrinogen oxidase) and the auxins (O) show tight  
 328 clustering in the optimal zone for phloem transport whilst also being mobile in the  
 329 xylem.



**Figure 5. Bromilow criteria by class.**

LogP vs. pKa clustering for selected classes of herbicides with the Bromilow regions overlaid to show that most herbicides do indeed adhere to the mobility rules. We observe specific clustering for individual herbicide classes. Mobility in phloem and xylem is indicated by the area labelled P & X; Mobility in Xylem = X; and immobile = I. OP = the region of the plot in which compound characteristics are optimal for phloem transport.

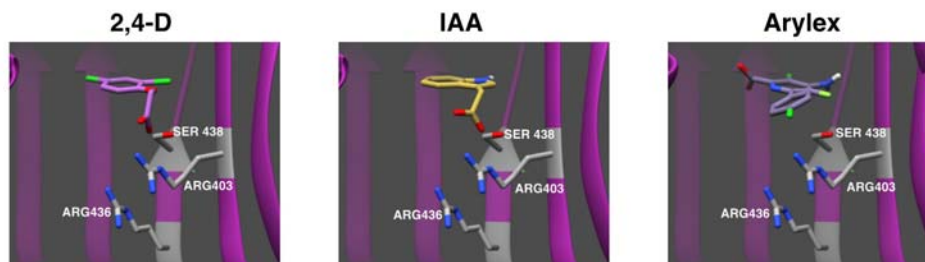
It is worth noting that many formulation-ready compounds, such as the esters noted above, would not be amenable for this type of analysis because they have no ionisable group. It is anticipated that the esters would be readily hydrolysed in the apoplast upon entry into plant tissue and the free acid would then pass around the plant systemically. This will be true for the auxins, and as free acids these all cluster tightly between LogP 0 to 3 and pKa's of 0 to 5 (Figure 5), similar values to those of IAA (logP = 1.71; pKa = 4.66; Supplementary Figure S7). Such tight clustering is not widespread, and class B has many compounds in the same pKa range but spread over a much wider range of LogP values, of -2 to 4.

Visualising the various distributions (Figure 1) and clusterings (Figures 4 and 5) by their HRAC class provides insight into class-specific properties. Previous cheminformatic reviews have tended to treat all herbicides collectively, and offered observations on the properties of herbicides on that basis. It is clear that compound

discovery programmes should consider the physiochemical properties for the appropriate target class. A summary of the properties of class O is given in Supplementary Figure 7.

#### **The properties of auxins and the structure of their target site:**

The auxin herbicides cluster tightly in all the druggability and transport metrics examined. One might imagine that all compounds targeted to specific sites would cluster together, although this is clearly not the case for many herbicide classes (Figure 5). This does suggest that the chemical space available for novel auxin discovery is confined, and we note that the natural auxin IAA sits within the cluster of synthetic analogues (Supplementary Figure S7 and Figure 4). We might also consider that a logical extrapolation of defining auxin-likeness is that these clustered molecules reflect the chemical nature of their binding site. This site, or more correctly this collection of sites, is well studied and the canonical representative is auxin receptor TIR1, for which the structure has been solved by crystallography (Tan et al. 2007). Auxins bind at the base of a deep binding pocket on TIR1, acting as “molecular glue” between the receptor and the degron peptide of Aux/IAA proteins. Further functional detail on the three-dimensional features of the TIR1 pocket have been revealed using novel tomographic docking tools (Uzunova et al. 2016). The structures of the auxin F-box proteins, AFB1-5, have not been solved, but a homology model for the member least like TIR1 (AFB5) has been presented (Calderon Villalobos et al. 2012). These auxin receptor proteins have also been used to develop a rapid biophysical binding assay using surface plasmon resonance (Calderon Villalobos et al. 2012; Quareshy et al. 2017) and a small library of auxin-similar compounds were screened to start to develop a pharmacophoric map for auxins (Lee et al. 2014). It is this map that can be anticipated to match and extend the definition of the cheminformatics parameters of auxins as herbicides. The SPR assay screens include many compounds that are not commercial herbicides and this wider screening will allow the development of a far more detailed set of rules for what features are detrimental, as well as features that are necessary for a molecule to be an active auxin.



**Figure 6. The auxin binding site.**

Crystal structure poses for 2,4-D and IAA in TIR1 (Protein Database repository codes 2P1P and 2P1Q respectively), with the carboxylic acid group adjacent to serine 436 and arginines 403 and 436. The aromatic rings adopt a similar pose. When Arylex™ is docked to TIR1 the results suggest poor binding given that no pose is identified that corresponds to that of IAA. This sub-group of picolinic acid-based auxins bind preferentially to AFB5 (Walsh et al., 2005).

In commercial terms, it is also relevant to consider the increasing problem of field resistance to auxin herbicides, and the question of whether or not we can identify more novel auxin herbicides, perhaps with different chemical scaffolds so that they are not susceptible to the same resistance mechanisms. In Arabidopsis, it has been shown that *tir1-1* mutant seedlings are clearly resistant to 2,4-D (Ruegger et al. 1998), whereas *afb5* mutants have been found resistant to synthetic picolinate auxins, with only minimal cross-resistance to 2,4-D or IAA (Walsh et al. 2006). These findings suggest that different auxin receptors can mediate chemical selectivity to different classes of auxin herbicides. All auxins are likely to conform to the properties of the class O physicochemical cluster, but we can envisage distinctive pharmacophoric maps associated with receptor sub-classes. To date there has been no record of field resistance to auxins associated with receptor site mutations (Heap, 2014), and it should be noted that field resistance is as likely to arise through changes to transport, metabolism and signal processing as it is through receptor mutation.

### Auxin discovery

Lipinski suggested that his “Rules” should be considered as guidelines, and cheminformatics is generally used as a mechanism for early screening of chemical compound libraries in order to reduce the burden of *in vivo* or *in vitro* screening experiments. To date, novel auxins have been discovered by directed chemistry and



414 structure-activity relationship screenings in vivo, combined with serendipity rather than  
415 such screening *in silico* (Epp et al. 2016). It is clear that this has been extremely  
416 effective given that 24 synthetic auxins (Supplementary Figure S7) have been used  
417 commercially to date, although not all remain available. With the molecular structure  
418 of the TIR1 receptor (Tan et al. 2007); Figure 6 rational design also became possible.  
419 Rational design is a strategy of creating new molecules with a defined functionality  
420 based upon the ability to predict how the molecule's structure will affect its behavior.  
421 This is generally an iterative process of prediction, test, learn and improve, and although  
422 it can be ab initio, it is normally done by making calculated variations and additions to  
423 a known chemical scaffold, a version of directed evolution. Rational design has already  
424 been used to deliver a set of anti-auxins to the research community (Hayashi et al.  
425 2012). Anti-auxins, such as one named as Auxinole, were built on the familiar auxin  
426 scaffolds of IAA, 2,4-D and 1-NAA, extending these into binding pocket space that  
427 would normally be filled with the degron of Aux/IAA proteins. Thus, the affinity and  
428 selectivity of known strong auxins were combined with the tools of molecular  
429 modelling, docking and synthetic chemistry to develop antagonist ligands.

430

431 Target site structures, therefore, offer alternative computational approaches for auxin  
432 discovery, including the use of docking software. As for all computational methods,  
433 docking is limited by the constraints of the system e.g. rigid protein structures. In TIR1,  
434 auxin binds at the base of a deep binding pocket and it has been shown that presenting  
435 the full volume of the pocket in a docking experiment will yield false positive results  
436 as well as masking actual binders (Uzunova et al., 2016).

437

438 Arylex and Rinskor belong to the picolinate class of auxins which have demonstrated  
439 selective binding to the AFB5 auxin receptor homologue (Lee et al. 2014; Walsh et al.  
440 2006). In the absence of the AFB5 crystal structure, the exact mechanism of binding  
441 for these AFB5-directed compounds is yet to be elucidated. If Arylex acid is docked  
442 into the crystal structure of TIR1 (Figure 6c) it fails to find a satisfactory pose. This is  
443 consistent with poor binding of picolines to TIR1 (Lee et al. 2014) but it also  
444 illustrates that rational design is an exacting science and it needs precise and detailed  
445 inputs to be helpful. Docking into a modelled AFB5 structure (Calderón Villalobos et  
446 al. 2012) is an option but, clearly, a model adds only another uncertainty into the

exercise. This is where alternative/orthogonal techniques such as cheminformatics could help with novel auxin discovery.

Tomographic docking examines constraints to the passage of the ligand into the pocket as well as final docked poses, and this does offer a more comprehensive protocol for ligand finding. Auxin was used as the exemplar ligand (Uzunova et al. 2016). However, tomographic docking is, necessarily, more onerous on computing power and result interpretation, and so it is valuable as a secondary screening method and primary screens are still needed to reduce chemical search space. For this, cheminformatics remains valuable, especially when the analysis is based on class “Rules”, as shown in Figures 2, 4 and Table 2.

We note that new synthetic auxins continue to be discovered and exploited (Epp et al., 2016), and have noted that the TIR1 receptor family is variously selective to different auxin chemical scaffolds. If the problem is inverted, we may ask whether or not we have discovered all natural auxins, and whether or not there is an endogenous analogue of the 6-arylpicolinate, for example. Combining physicochemical cluster rules with further receptor-ligand screening (Lee et al., 2012), the auxin pharmacophore, or pharmacophores, will become better resolved. Such advances may allow us to address what defines an auxin at atomic resolution, as well as help us discover new auxins.

### **Related discovery targets**

It is noted that rational design need not be limited to the ligand, and that receptor site design is also possible. One outcome might be herbicide tolerant crops, although other resistance mechanisms for auxin-tolerant crops are already in use (Behrens et al. 2007; Wright et al. 2010). The advent of these crops will, undoubtedly, increase auxin herbicide use and, hence, the need for more novel auxins.

One of the great values of auxin herbicides is their selectivity, generally lethal to dicot plants, whilst monocots are resistant, although e.g. quinclorac has activity against some grass weeds (Grossmann, 2010). The basis of this selectivity is still not fully understood and characteristics such as the anatomy of the vasculature, efficiency of detoxification, as well as of perception may all play a role. If it is receptor-based resistance, there is scope both for discovering additional compounds that break the

selectivity barrier and for using rational design to inverse the selectivity, giving grass-active auxin herbicides.

#### **Concluding observations:**

Global cheminformatics analyses of herbicides have been presented before, but we add to these a new database which has additional features for facilitating analysis, especially of selected classes of molecules. We have not discussed all the possible global trends, but have focussed on class O, the auxins, and have defined the known chemical space and features for these valuable compounds (Table 2). It is clear that analysis of the various herbicide property distributions (Figure 1) and clusterings (Figure 4) by their HRAC classes provides insights into the class-specific physicochemical properties and that these will help guide rational discovery programmes to maximise drugability and transport.

It is well known that the agrochemical world is facing the same challenges associated with the development of resistance as the pharmaceutical world is to antibiotic (antibacterial) resistances (Roca et al. 2015) The solution to this challenge is manifold, including improved stewardship (Mithila et al. 2011) but there is no doubt that molecule discovery will remain imperative, and for this we will need all the tools and resources available to us. We hope this work will serve as a useful resource for researchers working in herbicidal agrochemistry, although our DataWarrior database with its cheminformatics capabilities may also be helpful to those in the fields of fungicidal and insecticidal small molecule discovery.

Tables

HRAC Class	Pathway targeted	MW	H-bond acceptor count	H-bond donor count	cLogS	LogD	LogP	Polar surface area	Charges
A	Fatty acid biosynthesis	ns	ns	**	*	ns	**	ns	ns
B	Amino acid biosynthesis (Leu, Ile, Val)	****	****	*	ns	****	*	****	****
C	Photosynthesis (electron transfer)	****	ns	****	ns	ns	ns	*	***
E	Photosynthesis (heme biosynthesis for chlorophyll)	****	ns	***	****	ns	*	ns	ns
F	Photosynthesis (carotenoid biosynthesis)	ns	ns	ns	*	ns	ns	ns	ns
K1/K2	Microtubule polymerization	ns	ns	ns	ns	*	ns	ns	ns
K3	Fatty acid biosynthesis	ns	***	***	ns	***	ns	***	**
N	Fatty acid biosynthesis	***	****	**	ns	ns	ns	**	ns
O	Regulation of auxin-responsive genes	***	ns	ns	ns	***	ns	*	***
Z	Unknown	**	**	ns	ns	ns	ns	*	ns

**Table 1 Statistical analysis of individual herbicide class physiochemical features compared to all inclusive values**

Summary of one-way ANOVA of individual group means vs. the combined means for each herbicide class. ns = not significant i.e.  $P > 0.05$ , \* is for  $P \leq 0.05$ , \*\* is for  $P \leq 0.01$ , \*\*\* is for and  $P \leq 0.001$ , \*\*\*\* is for  $P \leq 0.0001$ . Statistical analysis were performed using GraphPad Prism version 7.0a for Max OS X. Multiple comparison statistical hypothesis testing was performed using the Dunnett methodology with a single pooled variance. Other general statistics used the Column statistics option.

	<u>Minimum</u>	<u>25% Percentile</u>	<u>Median</u>	<u>75% Percentile</u>	<u>Maximum</u>	<u>Mean</u>	<u>Std. Deviation</u>	<u>Std. Error of Mean</u>
MW	186.1	220	239.9	253.9	348	240.9	37.3	7.8
Hydrogen bond donor count	0	0	0	1	1	0.3	0.4	0.1
Hydrogen bond acceptor count	2	3	3	4	5	3.35	0.83	0.17
cLogS	-4.83	-3.81	-3.42	-3.03	-2.17	-3.40	0.60	0.12
LogD	-2.13	-1.12	-0.55	-0.24	0.14	-0.73	0.64	0.13
LogP	1.40	2.01	2.70	3.11	3.67	2.69	0.64	0.13
Polar Surface Area/Å <sup>2</sup>	37.3	46.53	46.53	76.21	85.44	56.6	16.5	3.4
Average Charge	-1	-1	-1	-1	-1	-1	0	0
Aromatic Ring count	1	1	1	1	2	1.2	0.4	0.1
Acidic pKa	0.89	2.41	3.1	3.6	4.75	3.0	0.9	0.2
Rotatable bound count	1	1	3	3	5	2.5	1.2	0.2
Total Surface Area/Å <sup>2</sup>	135.5	150.5	161.6	175.2	230.2	164.9	23.2	4.8

**Table 2. Physiochemical summary for synthetic auxins.**

## **A cheminformatics review of auxins as herbicides: References**

- Avram S, Funar-Timofei S, Borota A, Chennamaneni SR, Manchala AK, Muresan S.** 2014. "Quantitative estimation of pesticide-likeness for agrochemical discovery." *Journal of Cheminformatics* **6**, 1–11.
- Behrens M, Mutlu N, Chakraborty S, Dumitru R, Jiang WZ, et al.** 2007. Dicamba resistance: enlarging and preserving biotechnology-based weed management strategies. *Science* **316**, 1185-1188.
- Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL.** 2012. Quantifying the chemical beauty of drugs. *Nature Chemistry* **4**, 90–98.
- Bromilow RH, Chamberlain K, and Briggs GG.** 1986. Techniques for studying the uptake and translocation of pesticides in plants. *Aspects of Applied Biology* **11**, 29–44.
- Calderón Villalobos LIA, Lee S, De Oliveira C, Ivetac A, Brandt W, et al.** 2012. A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nature Chemical Biology* **8**, 477–85.
- Cobb A, Reade JPH.** 2010. *Herbicides and Plant Physiology*. 2nd ed. Wiley-Blackwell.
- Délye C, Jasieniuk M, Le Corre V.** 2013. Deciphering the evolution of herbicide resistance in weeds. *Trends in Genetics* **29**, 649–58.
- Dharmasiri N, Dharmasiri S, Estelle M.** 2005a. The F-Box protein TIR1 is an auxin receptor. *Nature* **435**, 441–45.
- Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, et al.** 2005b. Plant development is regulated by a family of auxin receptor f box proteins. *Developmental Cell* **9**, 109–19.
- Duke SO.** 2012. Why have no new herbicide modes of action appeared in recent years? *Pest Management Science*. **68**, 505-12.
- Epp JB, Alexander AL, Balko TW, Buysse AM, Brewster WK, et al.** 2016. The discovery of Arylex TM Active and Rinskor TM Active: two novel auxin herbicides. *Bioorganic & Medicinal Chemistry* **24**, 362–71.
- Gandy MN, Corral MG, Mylne JS, Stubbs K, et al.** 2015. An interactive database to explore herbicide physicochemical properties. *Organic & Biomolecular Chemistry* **13**, 5586–90.

- Ghanizadeh H, Harrington KC.** 2016. Cross-resistance to auxinic herbicides in dicamba-resistant *Chenopodium album*. *New Zealand Journal of Agricultural Research* **60**, 45–53.
- Goggin DE, Cawthray GR, Powles SB.** 2016. 2,4-D resistance in wild radish: reduced herbicide translocation via inhibition of cellular transport. *Journal of Experimental Botany* **67**, 3223–35.
- Grossmann K. 2010. Auxin herbicides: current status of mechanism and mode of action. *Pest Management Science* **66**, 113–20.
- Hayashi K, Neve J, Hirose M, Kuboki A, Shimada Y, Kepinski S, Nozaki H.** 2012. Rational design of an auxin antagonist of the SCF(TIR1) auxin receptor complex. *ACS Chemical Biology* **7**, 590–98.
- Heap I.** 2014. Global perspective of herbicide-resistant weeds. *Pest Management Science* **70**, 1306–15.
- Kepinski S, and Leyser HMO.** 2005. The Arabidopsis F-Box protein TIR1 is an auxin receptor. *Nature* **435**, 446–51.
- Knowles A.** 1998. Formulation of Agrochemicals. In: Knowles A. ed. *Chemistry and Technology of Agrochemical Formulations*. Springer Netherlands, 41–79.
- Lee S, Sundaram S, Armitage L, Evans JP, Hawkes T, Kepinski S Ferro N, Napier RM.** 2014. Defining binding efficiency and specificity of auxins for SCF TIR1/AFB-Aux/IAA co-receptor complex formation. *ACS Chemical Biology* **9**, 673–82.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ.** 1997. experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **23**, 3–25.
- Mithila J, Hall JC, Johnson WG, Kelley KB, Riechers DE.** 2011. Evolution of resistance to auxinic herbicides: Historical perspectives, mechanisms of resistance, and implications for broadleaf weed management in agronomic crops. *Weed Science* **59**, 445–57.
- Mockaitis K, Estelle M.** 2008. Auxin receptors and plant development: A new signaling paradigm. *Annual Review of Cell and Developmental Biology* **24**, 55–80.
- Oerke EC.** 2006. Crop losses to pests. *The Journal of Agricultural Science* **144**, 31–43.
- Quareshy M, Uzunova V, Prusinska J, Napier RM.** 2017. Assaying Auxin Receptor Activity Using SPR Assays with F-Box Proteins and Aux-IAA Degrons. In: *Plant hormones: Methods and protocols*, eds Kleine-Vehn J, Sauer M. Humana Press. **1497**, 159–191.

- Roca I, Akova M, Baquero F, Carlet J, Cavaleri M, Coenen S, et al.** 2015. The global threat of antimicrobial resistance: science for intervention. *New Microbes and New Infections* **6**, 22–29.
- Ruegger M, Dewey E, Gray WM, Hobbie L, Turner J, Estelle M.** 1998. The TIR1 protein of arabidopsis functions in auxin response and is related to human SKP2 and yeast Grr1p. *Genes and Development* **12**, 198–207.
- Sauer M, Robert S, Kleine-Vehn J.** 2013. Auxin: Simply complicated. *Journal of Experimental Botany* **64**, 2565–77.
- Schulz B, Segobye K.** 2016. 2,4-D transport and herbicide resistance in weeds. *Journal of Experimental Botany* **67**, 3177–79.
- Soltani N, Dille JA, Burke IC, Everman WJ, VanGessel MJ, Davis VM, Sikkema PH.** 2016. Potential corn yield losses from weeds in North America. *Weed Technology* **30**, 979–84.
- Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N.** 2007. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–45.
- Tice CM.** 2001. Selecting the Right Compounds for Screening: Does Lipinski's Rule of 5 for pharmaceuticals apply to agrochemicals? *Pest Management Science* **57**, 3–16.
- Uzunova VV, Quareshy M, del Genio CI, Napier RM.** 2016. Tomographic docking suggests the mechanism of auxin receptor TIR1 selectivity. *Open Biology* 160139.
- Walsh TA, Neal R, Merlo AO, Honma M, Hicks GR, Wolff K, Matsumura W, Davies JP.** 2006. Mutations in an auxin receptor homolog AFB5 and in SGT1b confer resistance to synthetic picolinate auxins and not to 2,4-Dichlorophenoxyacetic acid or indole-3-acetic acid in arabidopsis. *Plant Physiology* **142**, 542–52.
- Wright TR, Shan G, Walsh TA, Lira JM, Cui C, Song P, et al.** 2010. Robust crop resistance to broadleaf and grass herbicides provided by aryloxyalkanoate dioxygenase transgenes. *Proceedings of the National Academy of Sciences* **107**, 20240–45.



## Supplementary Information

Script for calculating physicochemical properties in DataWarrior

An sdf file containing 2D structures obtained from the HRAC website (<http://hracglobal.com/tools/classification-lookup> - last accessed December 2016) along with a unique compound number was created and used as the input sdf file in the script for calculating physicochemical properties.

The cxcalc routine was aliased in the bash.profile config, referencing the ChemAxon install folder to call up the binary, as required with the following code:

```
cxcalc -S -o <file path for output as .sdf> <file path to input sdf file with structures>  
<physiochemical property arguments here>
```

physiochemical property arguments:

- exactmass -p 3
- rotatablebondcount
- averagemicrospeciescharge -H 7.4
- acceptorcount -H 7.4
- donor -H 7.4
- name -t preferred
- aromaticringcount -a loose
- logd -H 7.4 --considertautomerization true
- logp
- pkacalculator -a 1 -b 0 -M true
- acceptor -H 7.4

The resultant sdf file for each property was imported into DataWarrior for analysing.

<b>HRA C class</b>	<b>Herbicide mode of action</b>	<b>Pathway or process targeted</b>	<b>No. of compounds</b>
A	Inhibition of acetyl-CoA carboxylase (ACCase)	Fatty acid biosynthesis	17
B	Inhibition of acetohydroxyacid synthase (AHAS, ALS)	Amino acid biosynthesis (Leu, Ile, Val)	55
C	Inhibition of photosystem II protein D1 (psbA)	Photosynthesis (electron transfer)	52
D	Diversion of the electrons transferred by the photosystem I ferredoxin (Fd)	Photosynthesis (electron transfer)	2
E	Inhibition of protoporphyrinogen oxidase (PPO)	Photosynthesis (heme biosynthesis for chlorophyll)	28
F	Inhibition of phytoene desaturase (PDS) or 4-hydroxyphenylpyruvate dioxygenase (4-HPPD) or of an unknown protein	Photosynthesis (carotenoid biosynthesis)	20
G	Inhibition of 5-enolpyruvylshikimate-3-phosphate synthase (EPSP synthase)	Amino acid biosynthesis (Phe, Trp, Tyr)	1
H	Inhibition of glutamine synthase	Amino acid biosynthesis (Gln)	2
I	Inhibition of dihydropteroate synthase	Tetrahydrofolate biosynthesis	1
K1/K2	Enhancement of tubulin depolymerization	Microtubule polymerization	19
K3	Inhibition of fatty acid synthase (FAS)	Fatty acid biosynthesis	26
L	Inhibition of cellulose-synthase	Cell wall biosynthesis	6
M	Uncoupling of oxidative phosphorylation	ATP biosynthesis	3
N	Inhibition of fatty acid elongase	Fatty acid biosynthesis	19
O	Stimulation of transport inhibitor response protein 1 (TIR1) + family of proteins	Regulation of auxin-responsive genes	24
P	Inhibition of auxin transport	Long-range hormone signaling	2
Z	Unknown	-	18

**Supplementary Table 1. HRAC classifications and their mode of actions – from (Délye et al. 2013).** The number of compounds in each class used in the current analysis is indicated in the final column.



Review of herbicides.dwar

File Edit Data Chemistry Database List Macro Help

Table	Structure No	Structure	Compound	Preferred IUPAC Name	HRAC class	Type	Molecular Weight
254	20		aminopyralid	4-amino-3,6-dichloropyridine-2-carboxylic acid	U	pyrimine carboxylic acids	200.302
255	28		chloramben	3-amino-2,5-dichlorobenzoic acid	O	Benzoic acids	204.970
256	28		dicamba	3,6-dichloro-2-methoxybenzoic acid	O	Benzoic acids	219.969
257	29		2,4-D	2-(2,4-dichlorophenoxy)acetic acid	O	Phenoxy-carboxylic acids	219.969
258	30		MCPA	2-(4-chloro-2-methylphenoxy)acetic acid	O	Phenoxy-carboxylic acids	200.024
259	31		picloram	4-amino-3,5,6-trichloropyridine-2-carboxylic acid	O	Pyridine carboxylic acids	239.926
260	36		dichlorprop	2-(2,4-dichlorophenoxy)propanoic acid	O	Phenoxy-carboxylic acids	233.985
	37		mecoprop	2-(2,4-dichlorophenoxy)propanoic acid	O	Phenoxy-carboxylic acids	214.040

Table Structures Binned Distributions Scatter Plots Counts Whisker plots

Selected: 1 Visible: 295 Total: 295

Category Browser

Column: HRAC class

Structure

contains

Structure No

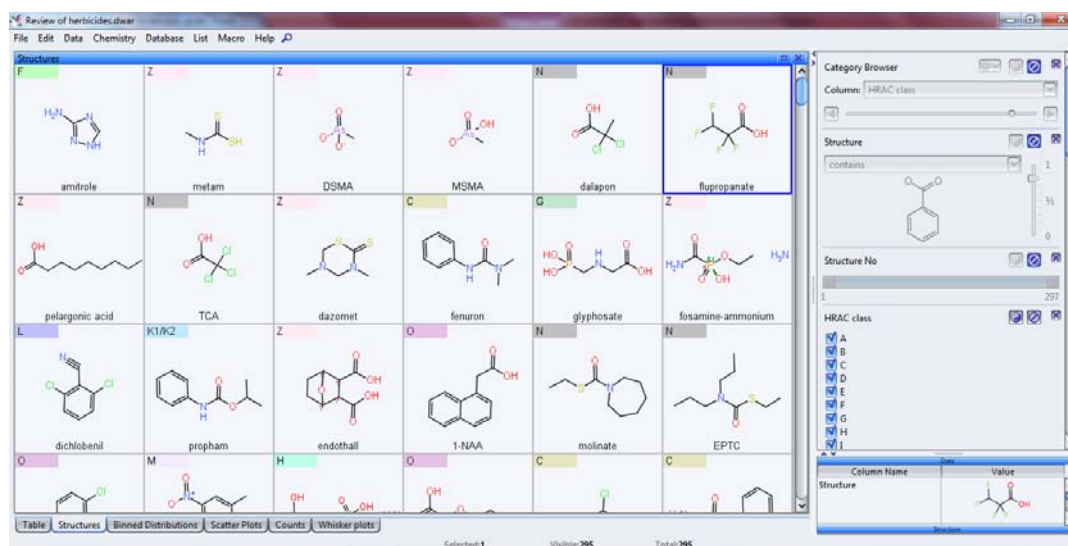
1

HRAC class

257

Column Name	Value
Structure	
Structure No	29
AROMATIC_RINGCOUNT	1
Molecular Weight	219.969
IUPAC name	2-(2,4-dichlorophenoxy)acetic acid
ROTATABLE_BOND_COUNT	3
AVERAGE_MICROSPECIES_CH	1
Donor count	0
Donor sites	0
Lead 2	1.03
Lead 3	1.04

**Supplementary Figure S1: A general overview the herbicide dataset in DataWarrior.** A tabulated view of chemical structures with 2,4-D highlighted in Red. The HRAC column has been coloured to give each group its own colour, as well as the Type of compound class. The Molecular Weight column is coloured as a rainbow gradient going from Red (smallest) to Violet (largest), although there is little variation amongst the auxins shown.

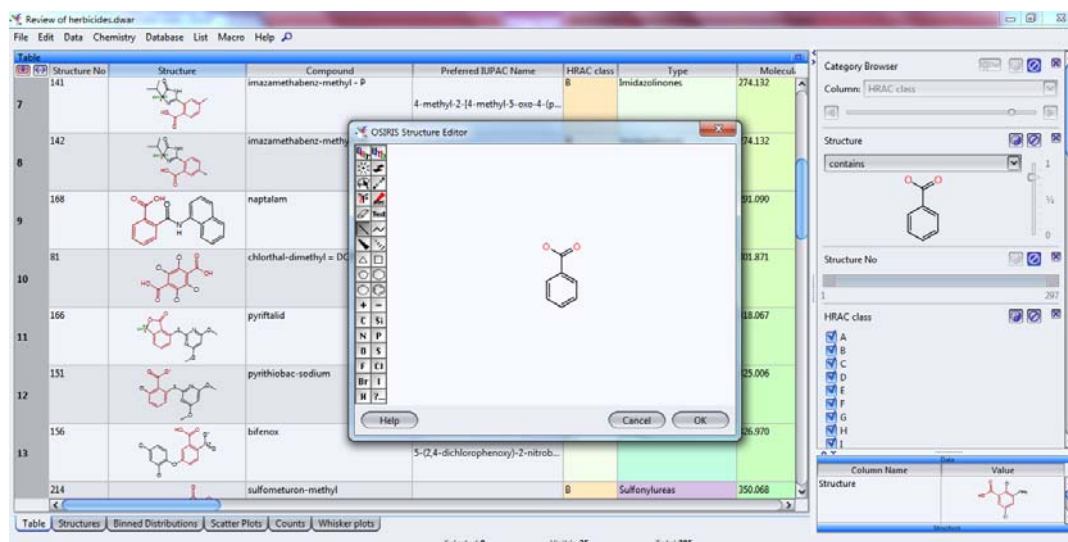


**Supplementary Figure S2: A structure-only view of compounds.** Sorted by Molecular weight, each compound HRAC class is displayed in the top left of the box, colour coded as in Supplementary Figure S1). The common name for each compound is displayed below its structure.

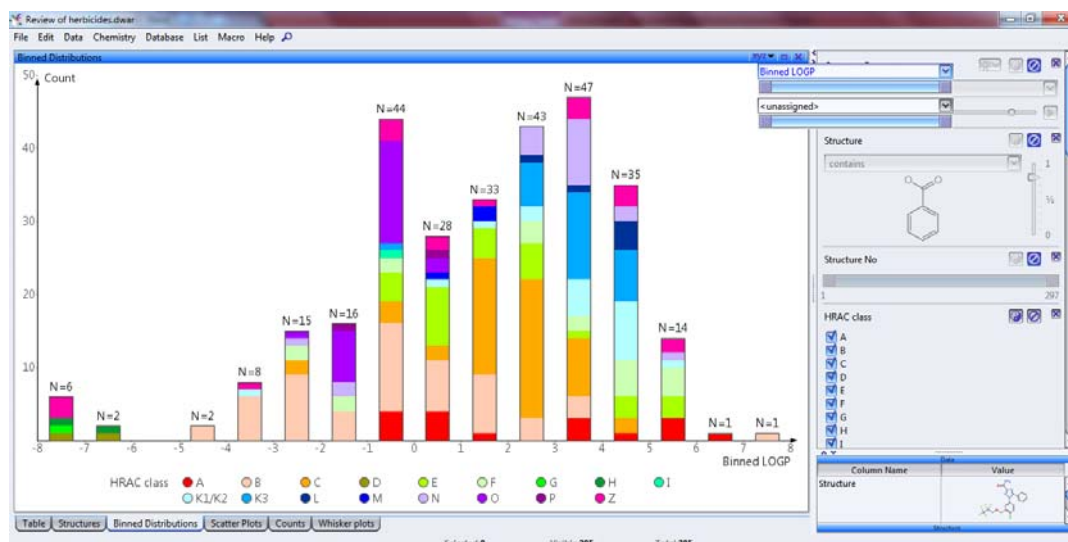
The screenshot displays the 'Review of herbicides.dbw' application. A central dialog box titled 'Calculate Compound Properties' is open, allowing users to select various physiochemical calculations for a chosen structure. The dialog has tabs for 'LE, Tox, Shape', 'Counts', 'Functional Groups', and 'Ionization'. The 'Counts' tab is currently active, showing a list of checkboxes for properties such as 'Druglikeness', 'Total average molecular weight in g/mol: natural abundance', 'Average molecular weight of largest fragment in g/mol: natural abundance', 'Monoisotopic mass of largest fragment in g/mol: most abundant isomers', 'cLogP: P: conc(octanol)/conc(water)', 'cLogS: % water solubility in mol/L, pH=7.5, 25C', 'cLogD at pH=7.4: via logP and ChemAxon pKa', 'H-Acceptors', 'H-Donors', 'Total Surface Area (SAS Approximation, Van der Waals radii, 1.4Å probe)', 'Relative Polar Surface Area (from polar and non-polar SAS Approximation)', 'Topological Polar Surface Area (TPSA, P, Ertl approach)', 'Druglikeness', and 'Permeability (cMDCK-PLS, F. Broccatelli, DOL10.1021/acs.molpharmaceut.6b00836)'. The background table lists herbicides with columns for Structure No., Structure, Compound, Preferred IUPAC Name, HRAC class, Type, and Molecular weight.

Structure No.	Structure	Compound	Preferred IUPAC Name	HRAC class	Type	Molecular weight
56		1-NAA	2-(naphthalen-1-yl)acetic acid	O	Aromatic carboxymethyl	186.068
14		clopyralid				190.954
30		MCPA				200.024
30		chloramben				204.970
18		aminopyralid				205.965
37		mecoprop				214.040
28		dicamba				219.969
29		2,4-D	2-(2,4-dichlorophenoxy)acetic acid	O	Phenoxy-carboxylic-acids	219.969

**Supplementary Figure S3: A dialogue box menu to select physiochemical calculations for structures in the database.**

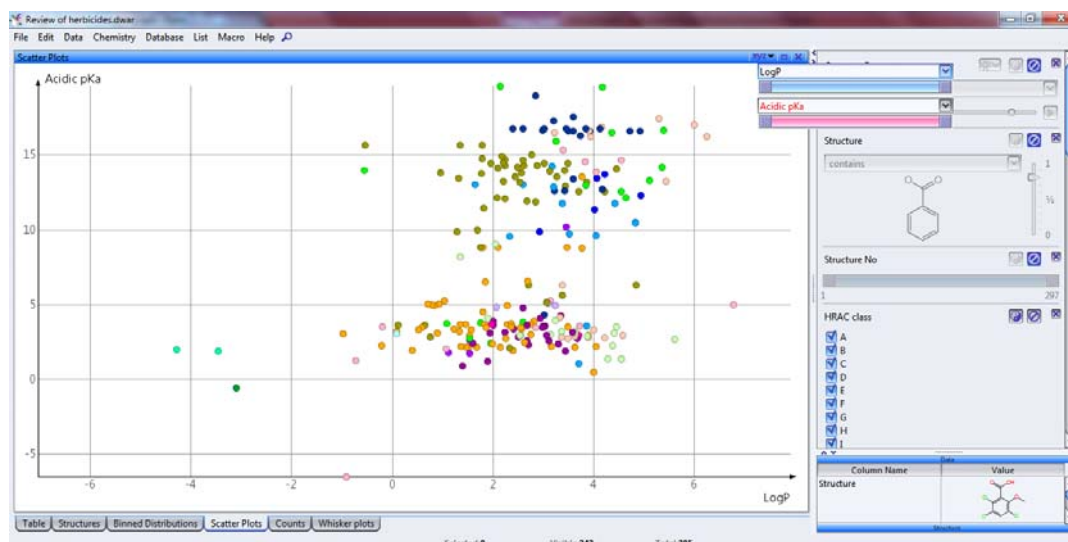


**Supplementary Figure S4: An example database search.** A search for compounds that have a benzoic acid moiety. The search yields 35 matches (displayed at the bottom of the window).




























**Supplementary Figure S5: An example graphical output.** The distribution of LogP values (binned) for all the HRAC groups by colour.





**Supplementary Figure S6: A scatter plot of LogP vs. pKa.** The axis can be changed on an ad hoc basis using the drop down selection menu on the top right hand side of the graph.

Structure	Compound	Molecular Weight	Donor count	Acceptor Count	cLogP	Acidic pKa	LogP	LogP	Ring Aromaticity	Charges	Electronegative Atoms	Total Surface Area	Polar Surface Area	Aromatic Rings
	IAA	173.187	1	2	-2.119	4.66	-0.96	1.71	2	-1	3	135.29	53.09	2
	1-NAA	186.068	0	2	-3.2	4.75	0	2.60	2	-1	2	145.72	37.3	2
	clopyralid	190.954	0	3	-2.893	3.15	-1.3	2.23	1	-1	5	126.96	50.19	1
	MCPA	200.024	0	3	-2.489	3.36	-1.12	2.41	3	-1	4	148.8	46.53	1
	chloramben	204.970	1	3	-3.177	3.6	-1.52	2.01	1	-1	5	136.72	63.32	1
	aminopyralid	205.965	1	4	-2.169	0.89	-2.13	1.40	1	-1	6	135.48	76.21	1
	mecoprop	214.040	0	3	-2.867	3.47	-0.55	2.88	3	-1	4	159.8	46.53	1
	dicamba	219.969	0	3	-3.119	2.54	-0.85	2.68	2	-1	5	150.46	46.53	1
	2,4-D	219.969	0	3	-2.881	2.81	-1.03	2.50	3	-1	5	151.96	46.53	1
	quinclorac	221.024	0	3	-3.415	3.63	-0.62	2.68	1	-1	4	158.4	50.19	2
	TBA	223.920	0	2	-3.837	1.9	-0.09	3.44	1	-1	5	143.62	37.3	1
	MCPB	228.055	0	3	-3.029	4.06	-0.59	2.94	5	-1	4	176.32	46.53	1
	dichlorprop	233.985	0	3	-3.259	2.95	-0.46	3.07	3	-1	5	162.96	46.53	1
	picloram	239.926	1	4	-2.905	3.9	-1.53	2.00	1	-1	7	150.9	76.21	1
	quinclorac	240.970	0	3	-3.807	3.57	-0.53	3.00	1	-1	5	161.56	50.19	2
	benazolin	242.976	0	4	-3.504	3.1	-1.59	1.94	2	-1	6	158.58	82.91	1
	clomeprop	248.001	0	3	-3.603	3.16	0.05	3.58	3	-1	5	175.22	46.53	1
	2,4-DB	248.001	0	3	-3.421	3.58	-0.5	3.03	5	-1	5	178.48	46.53	1
	Tricamba	253.930	0	3	-3.855	2.37	-0.24	3.29	2	-1	6	165.88	46.53	1
	2,4,5-T	253.930	0	3	-3.617	2.56	-0.42	3.11	3	-1	6	167.38	46.53	1
	fluroxypyr	253.966	1	5	-3.526	2.43	-1.94	1.59	3	-1	8	165.59	85.44	1
	triclopyr	254.926	0	4	-3.11	2.28	-0.83	2.70	3	-1	7	166.14	59.42	1
	Fenoprop	267.946	0	3	-3.995	2.7	0.14	3.67	3	-1	6	178.38	46.53	1
	Arylex	329.997	1	5	-4.516	3.18	-0.27	1.89	3	-1	8	223.85	85.44	2
	Rimskor	347.988	1	5	-4.83	4.24	-0.19	3.34	3	-1	9	230.2	85.44	2

**Supplementary Figure S7: Selected physiochemical data for auxins (HRAC class O).** The compounds are ranked by molecular weight.